Publication Bias in Clinical Research: Outcome of Projects Submitted to Ethics Committees

Jesse A. Berlin

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

In evaluations of medical or social strategies aimed at treatment or prevention of disease, the published scientific literature almost always plays a central role. Such evaluations may involve either informal reviews of the literature or, as is increasingly the case, formal research syntheses or meta-analyses. Regardless of which type of evaluation is undertaken, scientists and policy makers assume, often implicitly, that the published literature contains either all, or at least a representative sample of, the available data (1). The selective publication of scientific findings based on the magnitude and direction of the results—termed publication bias—could lead to serious errors in the estimation (from research syntheses) of the potential benefits of interventions. The purposes of this chapter are: (a) to present a brief description of selected early research on publication bias to provide a historical perspective; (b) to describe several empirical assessments of publication bias based on projects reviewed by institutional review boards (ethics committees); and (c) to present a brief discussion of some proposed solutions to the problem of publication bias.

PREVIOUS RESEARCH

The existence of publication bias has been a concern for about two decades in the social science literature (2–7). For example, White (4) reviewed research on the correlation between socioeconomic status and academic achievement and showed that the highest correlations are published in books (average correlation $r = 0.51$), compared with journal articles ($r = 0.34$) and unpublished material ($r = 0.24$). Glass et al. (6) found systematic differences in the magnitude of effects between journal articles and dissertations in a variety of meta-analyses in psychology. Smith (5), in summarizing studies of educational innovations, found that effect sizes were considerably smaller for unpublished studies than for published studies. Coursol and Wagner (7) surveyed psychological researchers who had conducted studies of therapeutic outcome. They found that both the decision to submit an article for
publication and the editorial decision to accept or reject were associated with the results of the research, with statistically significant results favored at each step. Specifically, they found that 82% of articles with a significant outcome were submitted, compared with 43% for articles with a nonsignificant outcome. Among articles submitted, 80% of the significant articles were accepted versus 50% of the nonsignificant studies. A consequence of these findings is that inferences based on published literature may be biased away from the null hypothesis.

The early work on publication bias has led to increasing concern about its existence in the medical literature and evidence that it is a problem (8–12). One early approach to assessing publication bias used only published literature by exploiting the relationship between sample size and effect size (8–10). The authors argued that in order for studies with small sample sizes to achieve statistically significant results, those studies would have to show large effect sizes. Thus, if publication bias exists, one would expect to see large effect sizes, on average, in small studies, and smaller effect sizes in larger studies. Working with a consecutive series of published reports of cancer therapy, Begg and Berlin (8,9) found strong evidence of the anticipated relationship between sample size and effect size. In subsequent work, Berlin et al. (10) found that the apparent magnitude of publication bias, as evidenced by the strength of the association between sample size and effect size, depended on several study characteristics. A finding of particular interest was that randomized trials showed little or no evidence of publication bias; that is, there was only a weak association between sample size and effect size, whereas nonrandomized (but still comparative) trials showed considerable bias (10).

In some clinical fields, it is possible to use a sampling frame that is unaffected by publication bias. Simes (11,12) examined studies that compared single alkylating agent therapy with combination chemotherapy in the treatment of advanced ovarian cancer. He restricted the summary to randomized trials. Using several strategies, he found 20 published trials. He found an additional six studies through a data base maintained by the National Cancer Institute, the Compilation of Experimental Cancer Therapy Protocol Summaries (13). Three of the four significant trials were published but not registered, whereas all six unpublished trials yielded nonsignificant results. When Simes performed a standard meta-analysis using only published studies, he observed a significant improvement in survival with use of combination chemotherapy \( (p = 0.02) \). When the analysis was restricted to registered studies, for which data were available in 13 of 14, the observed improvement in survival was smaller, and the summary result was no longer significant \( (p = 0.24) \).

Following these studies, several empirical assessments of publication bias of a somewhat different nature were conducted (14–16). These four studies, described in three papers, involved the identification of projects that were submitted to institutional review boards (ethics committees) and follow-up of those projects to determine whether they had been published or not and whether they had obtained statistically significant findings. The design and results of these studies are the focus of the next section of this paper.
EMPIRICAL ASSESSMENTS OF PUBLICATION BIAS

The Oxford Study

The Oxford Study of publication bias was initiated in May 1990 and was reported in *Lancet* (14). The files containing all research protocols submitted to the Central Oxford Research Ethics Committee (COREC) between January 1, 1984, and December 31, 1987, provided the framework for this study. From those files, the investigators abstracted protocol titles and the names of the principal investigators. They wrote to all of the principal investigators explaining the purpose of the study and followed up with a telephone interview for information on the status of the study as of May 1990.

For projects that had actually begun, further information on the design, organization, results, and publication status of each study was obtained in the interview. Coinvestigators were contacted in the absence of, or at the request of, the principal investigator. The information requested included the source of funding of the study, the final sample size, the nature of the comparison group if one existed, and the main study findings, especially the attainment of statistical significance. Further information was requested regarding publications and presentations, papers rejected, planned, or still under review, and reasons for nonpublication. For clinical trials, a specific question was asked about the use of randomization.

The important specific definitions were as follows:

1. An experimental study was defined as one in which one or more variables were controlled by the investigator in order to study the effect on an outcome measure. This included clinical trials, in which a particular procedure or treatment was being studied, usually in comparison with another treatment or procedure.

2. Studies were considered statistically significant if the main outcome achieved a p value of 0.05 or less, as showing a nonsignificant trend if the difference or association carried a p value greater than 0.05, or as null if no difference was observed between the study groups or if no significant correlations were observed.

3. Publication implied acceptance by a journal but did not include book chapters or published meeting abstracts or proceedings.

The principal statistical analysis involved the use of logistic regression to examine predictors of the dichotomous outcome variable representing publication status. Separate analyses were conducted using the outcome “published or presented” and yielded similar results. A main question of interest was whether the magnitude of publication bias varied across different subgroups of studies, so appropriate interaction terms involving the “statistical significance” variable and each of several covariates were also examined.

The investigators identified 720 studies approved by the COREC during the 4-year period. Studies were excluded from further consideration if they had been subsequently withdrawn by the original investigators (n = 4); if they were not formal research studies (n = 5); if they were lost to follow-up because the principal
investigator could not be contacted \((n = 172)\); because the original investigators provided inadequate information \((n = 15)\); or because the investigators failed to respond to the mailed questionnaire \((n = 28)\), leaving 487 studies contributed by 216 investigators. Comparisons between those studies for which the investigator was found and interviewed and those for which the investigator was not located or did not respond revealed no significant differences in the year approved. The number of studies approved by the COREC per investigator, the main department of the study, or the type of study design.

As of May 1990, 287 studies had been partly (recruitment alone) or fully (recruitment and follow-up) completed; 100 had never started; 58 had been prematurely terminated or suspended; and 42 were still in progress. Only those studies that had been analyzed were considered to have the potential for being written up and published, so tests for publication bias were restricted to the 285 studies that had been analyzed, including 148 clinical trials.

By May 1990, 138 (48\%) of the studies had been published. The unadjusted relationship between significance of the main outcome and publication status showed strong evidence of publication bias. Sixty-seven percent of published studies had statistically significant findings, compared with only 29\% of studies neither published nor presented. Only 15\% (23/154) of studies with significant results remained unpublished or unreported compared with 44\% (43/97) of those with null results. The unadjusted odds ratio (OR) for publication with a significant result versus a null result was 2.96 (95\% confidence interval, CI: 1.68 to 5.21), indicating a higher likelihood of publication with a significant than with a null result. For the 148 clinical trials alone, the association was smaller, with an unadjusted OR of 2.10 (CI: 0.98 to 4.52). After adjustment for multiple other factors using logistic regression, the odds ratios for significant versus null studies were 2.32 (CI: 1.25 to 4.28) for all studies and 1.59 (CI: 0.70 to 3.60) for the clinical trials.

In the logistic regression analyses, other predictors of publication included sample size \(>20\) (OR = 1.74; CI: 0.95 to 3.18) and funding source. Specifically, industry-sponsored studies were significantly less likely to be published (OR = 0.36; CI: 0.16 to 0.83).

Of the 78 unpublished studies, 23 (29\%) had significant results. The most frequent reason given by the investigators for not publishing was that a paper had been written but not yet submitted, or submitted but not yet accepted \((n = 35)\) of 175 reasons, 19\%). For the 43 unpublished studies with null results, the presence of the null result was the most frequent reason for failing to write up the study. Editorial rejection was cited infrequently (9\%) as the reason for studies remaining unpublished.

The authors also examined the association between the statistical significance of a study and the likelihood of it being submitted for publication. The adjusted OR of 2.94 (CI: 1.43 to 6.01) for submission was higher than the OR of 2.32 for publication, suggesting that the investigators of the original studies, rather than journal editors, play a major role in publication bias.

Differences in the degree of publication bias were examined between various subgroups of studies by testing the covariate statistical significance interaction terms.
in the logistic regression models. The only significant difference in the degree of publication bias between subgroups was related to study design. The adjusted OR for publication bias in observational and laboratory-based studies (including both comparative and noncomparative studies) was 3.79 (CI: 1.47 to 9.76), compared to only 0.84 (CI: 0.34 to 2.09) for randomized clinical trials. For clinical trials alone, randomized trials were significantly less susceptible to publication bias (OR = 0.73; CI: 0.28 to 1.91) than nonrandomized (but still comparative) trials (OR = 10.26; CI: 1.76 to 60).

Johns Hopkins Studies

Because the design of these investigations was so similar to that of the Oxford Study described above, I focus mostly on comparisons between the approaches. These studies were performed by Dickersin and colleagues at Johns Hopkins University (15). The original studies forming the basis for this research were those appearing on the records of the two institutional review boards serving the Johns Hopkins Health Institutions and approved in or before 1980. One board served several medical institutions (to be referred to as MED), and the other served the School of Hygiene and Public Health (referred to as PH). The logs of the two boards contained 1048 applications (MED, 766; PH, 282). Trained project staff were responsible for locating logged applications in files, for abstracting and recording information from the applications, and for classifying the study designs. The applications were read independently by two readers, and disagreements were adjudicated by a master reader.

Three hundred eleven applications were excluded because they were either withdrawn or not approved (MED, 28), were approved but not implemented (MED, 133; PH, 51), were classified as exempt from review (MED, 23), did not describe a research study (for example, applications for training) (MED, 26; PH, 6), or did not involve humans (MED, 19; PH, 25). The remaining studies were considered eligible for interview.

In 1988, the principal investigators of the 737 interview-eligible studies, or their surrogates, were contacted by trained interviewers in random order. The nature of the study findings was classified as in the Oxford Study (14) (statistically significant, suggestive trend but not statistically significant, or no trend or difference); however, for analytic purposes, significant studies here were contrasted with the combined group of trend and nonsignificant studies. (In the Oxford Study, significant studies were contrasted directly with the null studies.) Studies were classified as published if they were reported in journal articles, books or book chapters, monographs, or were available from medical libraries or from a public archive (for example, the National Technical Information Service). These authors did not examine publication in a journal as a separate outcome.

The results of the Hopkins studies were very similar to those of the Oxford Study (14). At the time of the interview, 277 (81%) of the 342 MED studies had been published, as had 113 (66%) of the PH studies. For MED, 89% of the significant
studies had been published, compared with 69% of the nonsignificant studies (OR = 3.38; CI: 1.96 to 5.83). For PH, the values were 71% and 58%, respectively (OR = 1.78; CI: 0.94 to 3.39). Several other variables were predictive of publication status, including funding source. For both MED and PH, industry-funded studies had the lowest publication rates (65% and 50%, respectively), but the significance of that unadjusted result was not maintained in multiple regression analyses. Sample size (≥100 versus <100) was not significantly associated with publication status in either MED or PH in logistic regression models.

The reasons for not publishing were similar in these studies to those in the Oxford Study (14). Over 90% (MED, 97%; PH, 93%) of the unpublished studies were not published because of the actions or inactions of the original investigators. Only a small proportion (MED, 3.1%; PH, 6.8%) of papers remained unpublished because they had been rejected by a journal.

National Institutes of Health Clinical Trials

The method for identifying eligible studies was different in this investigation from in the other two, but the strategy was similar. This study was also conducted by Dickersin and her colleagues at the Johns Hopkins University (16). Studies were identified from the magnetic tapes of the 1979 Inventory of Clinical Trials from the National Institutes of Health (17). A total of 986 trials were available on the tapes, 654 funded by the National Cancer Institute (NCI) and 332 by other institutes. The NCI studies were excluded because they were thought to include a large proportion of "trials" that were not stand-alone trials but were parts of ongoing programs. As in the other studies described above, principal investigators or their surrogates were contacted and interviewed about study characteristics and outcomes. Definitions of variables were essentially the same as those used in the Johns Hopkins studies.

Of 332 trials listed on the tapes, 293 (83%) were eligible for interview. Complete interviews were obtained for 217 (74%) of the trials. The 198 trials for which there was publication information form the basis of the analysis. These studies were found to be comparable on several dimensions to the studies for which incomplete or no information was available. A large proportion (184/198, 93%) of the trials had been published. Publication was more likely for trials reporting a statistically significant finding than for those with a nonsignificant finding (OR = 7.04; CI: 1.90 to 26). No other trial or author characteristic was associated with publication status. However, publication bias appeared to be absent in multicenter trials (OR = 0.84; CI: 0.07 to 9.68) but severe in single-center trials (OR = 21; CI: 2.60 to 172).

Meta-Analysis of All Four Studies of Publication Bias

Dickersin and colleagues (16) performed a meta-analysis of the four studies described above. The results are summarized in Fig. 1. The combined, unadjusted OR
was 2.88 (CI: 2.13 to 3.89). There was little evidence of heterogeneity of the odds ratios across studies of publication bias.

**SOLUTIONS TO THE PROBLEM OF PUBLICATION BIAS**

The existence of publication bias is clearly demonstrated by the studies described. Thus, the potential for obtaining biased summaries from research syntheses based solely on published studies is real and of concern. Various methods have been proposed to estimate the potential magnitude of publication bias in the context of performing a meta-analysis, to correct for the bias analytically, and to prevent its occurrence. I describe each of the approaches briefly in this section.

A simple and practical technique for assessing the potential for publication bias is the “funnel plot,” first proposed by Light and Pillemer (18). This method involves plotting the measure of effect (for example, the log odds ratio) against a measure of study size, either the actual sample size or the inverse of the variance of the effect measure. In the absence of publication bias, the points should produce a funnel shape such that the values of the effect size are scattered symmetrically around the true estimate, with that scattering narrowing as the sample size increases. Figure 2 (19) shows an idealized funnel plot. If publication bias exists, there may be few or no points around the point estimate, indicating no effect (for example, a log odds ratio of 0) for the small studies. Alternatively, points may appear to be missing on one side of the plot.
Fig. 2. Idealized funnel plot of expected scatter of study results according to study size. $X_T$ indicates a true positive effect; 0 indicates a null effect; and "Gap" indicates the expected lack of published results in the event of publication bias (19).

Figures 3 and 4 show two funnel plots representing studies of psychoeducational programs for surgical patients (18,20). In the first plot, only the published studies are included, and the funnel appears to have a gap where the small studies showing no effect of these programs should be. The second plot includes unpublished studies, and the former gap is filled by those studies.

A recent method proposed to assess publication bias involves the use of a modified version of the ordinary Kendall’s $\tau$ rank correlation procedure (21). It is a direct statistical analog of the funnel plot described above. Although this test is based on the appealing analogy with the funnel plot and is relatively free from restrictive assumptions, it is not very powerful statistically when the number of studies is 25 or fewer.

An early approach to quantifying publication bias in a given data set was developed by Rosenthal (22). He proposed the calculation of a quantity he called the "fail-safe $N" when the results of a meta-analysis are statistically significant. His method uses the Z statistics from the individual studies included in the meta-analysis to calculate the number of unpublished studies showing no effect—that is, with a $Z$ statistic of zero—that would have to exist to render the overall combined $Z$ statistic
FIG. 3. Funnel plot for published studies only: analysis of data from Devine and Cook's (20) review of psychoeducational programs for surgical patients. (From Light and Pillemer, ref. 18, reprinted by permission of the publishers. Copyright 1984 by the President and Fellows of Harvard College.)

FIG. 4. Funnel plot for published (■) and unpublished (▲) studies combined: analysis of data from Devine and Cook's (20) review of psychoeducational programs for surgical patients. (From Light and Pillemer, ref. 18, reprinted by permission of the publishers. Copyright 1984 by the President and Fellows of Harvard College.)
(including both published and unpublished studies) nonsignificant. If the number of unpublished null studies required to overturn the significant result is small, then the meta-analysis should be interpreted with some caution. If that number is large, then the conclusions of the meta-analysis are unlikely to change even if a few unpublished studies do exist. In general, this approach is of limited utility for two reasons: first, because it uses only Z statistics and ignores quantitative estimates of effects (for example, odds ratios); and second, because the assumption that all the unpublished studies have a Z statistic of exactly zero is unrealistic.

Methods of correcting for publication bias deal specifically with the issue of estimating bias-corrected effect sizes but make some fairly strong assumptions about the underlying mechanisms leading to the bias. For example, a method proposed by Iyengar and Greenhouse (23) examined two particular functional forms relating the p value attained by a study to its probability of publication. They assume in one approach, for example, that the probability of publication is proportional to the t statistic for each study when the study is not statistically significant and that all statistically significant studies are published. The available data are used to estimate both the average effect size—corrected for publication bias using weighted distributions—and the probabilities of publication. A similar but more recent approach relaxes some of the assumptions about the selection mechanism (24). This method still assumes that the p value is the primary determinant of publication; however, it does not assume a particular shape for the relationship between the p value and the probability of publication. The authors propose this method as an exploratory technique, intended as an informal tool to assist in establishing the likelihood that publication bias is a serious problem in a given meta-analysis. They suggest, however, that when bias is identified by their approach, one should be very cautious about using their model (or any other) to correct it. They argue that attention should instead be directed toward identifying the causes of the bias, perhaps by initiating a search for missing studies (24).

Because none of the available methods is entirely satisfactory for correcting publication bias, efforts should be directed at its prevention. One way to avoid publication bias would be to obtain unpublished studies. In some cases, unpublished data can represent a large proportion of the available data. For example, in the Early Breast Cancer Trialists Collaborative Group meta-analysis (25), 20 of the 92 studies (22%) providing data for the meta-analysis were unpublished. It is often impracticable, however, to identify unpublished studies (26). Further, some investigators have chosen not to include unpublished studies in a meta-analysis because those studies have not usually been peer-reviewed (27). A survey of authors of meta-analyses, meta-analysis methodologists, and editors showed that most (78%) of the meta-analysts and methodologists favored the inclusion of unpublished material in meta-analyses under some conditions, whereas only 47% of the editors favored the inclusion of unpublished data (28).

In the clinical trials field, there has been some movement toward the prospective registration of all initiated studies (29–33). This would provide a sampling frame that would identify studies independently of their findings, much along the lines of
the Simes (11,12) study described above. Funding for such registries has been
difficult to obtain and seems to remain an issue. Various institutes within the National
Institutes of Health support registries of studies involving human beings (31,32). The
systems of ethics committees and institutional review boards in place in many
countries could provide a convenient framework for such registration.

CONCLUSIONS

The empirical studies of publication bias, coupled with previous research, show
that publication bias is a potentially serious problem when a meta-analysis is per-
duced or even when individual studies are evaluated. Although there is some evi-
dence that publication bias is not as severe for randomized trials as for nonran-
donized trials, the potential for bias to exist should not be dismissed solely on the
basis of the presence of randomization. In any meta-analysis, funnel plots or related
analytic approaches should be used to evaluate the potential for publication bias. If
there is evidence of publication bias, results of the meta-analysis need to be inter-
preted extremely cautiously. Because it is not clear that current analytic corrections
for publication bias are entirely satisfactory, the best solution to publication bias is
prevention. One prevention strategy is to try to obtain unpublished studies, but this
is not always feasible without registries of trials and is not universally viewed as
desirable, given the lack of peer review for unpublished studies. Registries—perhaps
using the framework of existing ethics committees or institutional review
boards—represent a promising strategy.

REFERENCES

1. Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: the effect of the outcome
2. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of
4. White KR. The relation between socioeconomic status and academic achievement. Psychol Bull
   1982;91:461–481.
   tions, 1981.
7. Coursol A, Wagner EE. Effect of positive findings on submission and acceptance rates: a note on
   meta-analysis bias. Profess Psychol 1986;17:136–137.
   151(part 3):419–463.
    4:1529–1541.
    summaries. NIH Publication No 83-1116. 7th ed. Washington, DC: US Department of Health and
    Human Services, 1983.
DISCUSSION

Dr. Aeschlimann: Because of the bias toward publishing positive results, I am afraid that if I want to do a meta-analysis on safety data from published results, the final result will be unreliable.

Dr. Berlin: Safety issues are also complicated by the fact that we are often looking at very small event rates, particularly with adverse effects of drugs, and those are very hard to detect in individual studies. So you are dealing with an added complication that the studies, in general, are going to be too small to look at the issue properly, and then, there is also a zero event rate in both groups, which is quite often the case in some of the safety studies. One suggestion is that we really need larger safety studies to begin with.

Dr. Guesry: There may be an additional publication bias because of the problem of publication in little known journals. I suspect that the visibility of the journal in which the study is
published is very different depending on whether the results are significant or nonsignificant. Some of the thousands of publications are barely available in libraries.

Dr. Berlin: Dickersin looked at that in the Johns Hopkins study, and your suspicion was correct: the significant studies were more likely to be published in a high-visibility journal. A Scandinavian study (1) also looked at the issue of reference bias: when you look at a review article that summarizes the existing literature, and you examine the references that are cited by that article, and then do your own search to find all of the existing literature, then there seems to be preferential citation of the significant results. I think part of the issue here is that scientists need to get promoted; and to get promoted, you have to publish; and to publish, you have to find significant results—it’s a vicious cycle. One point I didn’t make clearly enough is the focus on p values; if we had started out looking at likelihood methods or even confidence intervals, then we might have avoided a lot of publication bias—it is the focus on p values and p < 0.05 that is really the issue here.

Dr. van’t Hof: Although I am convinced of the existence of publication bias, I wonder if you have overestimated it because the quality of the study itself may be a confounder in this situation. If the quality of the study is bad, the results will not be good, or there will be no useful result, so in this case, you hesitate to publish.

Dr. Berlin: I think you are asking whether statistical significance, publication status, and study quality are confounding variables. Even after you adjust for various measures of study quality, you still see this publication bias effect. I think the more interesting question is probably the interaction question; that is, are poor-quality studies subject to a different degree of publication bias than good-quality studies? What we are suggesting, and are not really able to prove from one example, is that higher-quality studies and particularly randomized trials seem to be less prone to publication bias than others. They are not immune from publication bias—in fact, you can find examples of groups of randomized trials where there is some evidence of publication bias.

Dr. Madden: The idea of an electronic journal is intriguing. Would it be peer-reviewed? Do you see this as a trend? And also, could you say something about your thoughts on a registry?

Dr. Berlin: Henry Sacks is the editor of the one journal that I know about, which is the On Line Journal of Current Clinical Trials. It is to my knowledge the only existing electronically published journal. It is peer-reviewed. There was discussion ad nauseam when we were setting the journal up about the issue of peer review, and let me reassure you that if anything, we are erring on the side of too much peer review because of potential criticism. There is a risk with any kind of electronic medium that proper peer review will not be implemented. As to the registry issue, I think the idea of a registry is fairly widely accepted in the clinical trials community in principle, but the problem has been that nobody wants to pay for a registry, and that is part of the impetus for the Cochrane collaboration. What is happening is that people in the Cochrane collaboration are in effect creating their own registries, at least of published trials, using their own resources and their own time. This is something that we are moving toward, whether or not anybody wants to pay for it.

Dr. Uauy: I would encourage the model of the Cochrane data base, where people register their protocols before they have results, and at the present time, they are providing data from all of the registered studies, whether they are published or not. I think that is going to be the only way of addressing this problem in the long run. The other point I would like to make is that in the literature, and I think this model is true for a lot of pediatric interventions, there are usually several small to medium-sized trials showing effects, and then you do a collaborative prospective randomized trial, and you have no effect; we have at least three or four in the
neonatal literature over the last few years—intravenous immunoglobulin, comparisons of corticosteroids for the prevention of various conditions, and so on—and we find that the large control trials show no effects. So I think that the publication bias in favor of positive results is affecting clinical practice in a deleterious way, and we should be more demanding about what we accept as necessary evidence to modify clinical practice, whether it is published or not.

**Dr. Berlin:** One point I didn’t make clear enough is that when I spoke about registration, I was really talking about prospective registration, that is, registration before the conduct of the study—or before you know what the answer is, you register the trial, so that; even if you never finish the trial, somebody knows that it existed or should have existed. That is the only way of avoiding any kind of selection bias in the registry itself. I think you are touching on an issue about the fundamental process of how medical research takes place. In fact, it is not a random process at all—the statistical models all have some assumptions about random sampling. The kind of process that you are referring to reflects the fact that it is the positivity or possibly the negativity of preceding research that stimulates new studies. The whole statistical framework of the null hypothesis is bogus in that context, because nobody would really undertake a study based on the belief that the null hypothesis was true.

There is a statistical framework that says, “here is the null,” but what I am really trying to do is to knock down the null hypothesis. However, if I didn’t have some reasonable prior suspicion that the null is not true, why would I bother spending all that money and all that time? I think the situation relating to how studies get done is probably more chaotic than we like to believe, and it may be an argument for prospective meta-analysis. One reason why the FDA has not gotten more involved in meta-analysis as a way of evaluating the evidence is that it is a retrospective process, and we don’t know why the research got done or how it got done. In fact, what we should be doing is sitting down right at the beginning and saying here are the clinical issues, here are the subgroups that are important, so what is the best way to design not the first trial but the next five trials that you are going to need to provide the evidence that will convince us one way or the other that the treatment works or doesn’t work?

**Dr. Clarke:** The title of this chapter was “Publication bias in clinical research,” and yet, you have dealt almost exclusively with clinical trials, and of course that doesn’t cover the full range of clinical research, which also makes a very helpful contribution in straightforward observational clinical work and sometimes in retrospective assessment of clinical data and that sort of thing—it doesn’t always have to be trials. I am the most avid supporter of everything you should do about trials—randomizing, double-blinding, etc.—but I do sound a slight note of caution. For those of us who have to make policy decisions on the available data, sometimes the controlled trials don’t come up with the best answers, so far as one can assess. If you take, for instance, folic acid supplementation in the prevention of neural tube defect, the only prospective controlled trial comes up with a value of 800 μg of folic acid per day, whereas I think universally now the value advised is 400 μg, and that is based almost entirely on retrospective data. I wonder if you could comment?

**Dr. Berlin:** I may have emphasized the clinical trial aspect of publication bias studies, but in fact, the Oxford Study covered the entire range of studies. In terms of retrospective studies, I think that if you get into retrospective versus randomized studies in a given clinical area, there are going to be strengths and weaknesses to each. In breast cancer screening, for example, there are some obvious strengths to the large randomized trials, but there are some obvious weaknesses also in terms of compliance and crossover and contamination of groups. So I would not dream of saying that the only thing we should base our clinical decisions on is randomized trials—clearly, in some cases, it is not feasible to do randomized trials. I think
the point that you are making is that somebody has to make a decision on the basis of the data, and those data may be better in some cases than in others.

*Dr. Rey:* The story of folic acid and neural tube defects is more complicated than Dr. Clarke stated. First of all was the MRC study, the secondary prevention study in women who had already given birth to a child with a neural tube defect. In this, they used 4 or 5 mg a day. Then there was the Hungarian study, a primary prevention study using 700 µg/day with other vitamins added, and it is not clear to this day whether folic acid alone is sufficient to prevent neural tube defects in primary prevention. It is also amusing to remind ourselves that an ethics committee at the beginning of the 1960s rejected a double-blind versus placebo study, which was finally realized by the MRC 10 years later. So the actions of an ethics committee may also be a possible source of bias! I know another story of bias introduced by the peer-review system: the original research of Weijers and van de Kamer on the role of gluten in celiac disease was rejected by an American journal because, in the opinion of the reviewer, it was impossible that bread could be responsible for celiac disease. This is the reason why they published all their papers, except one, in *Acta Paediatrica Scandinavica.*

*Dr. Roche:* Could you clarify what you mean when you ask a principal investigator whether he has published the results of his study. Many studies are multifaceted, and results could easily be published that are not central to the major hypothesis addressed by the study. He could still say he published the results, even if this was only a description of the recruitment process.

*Dr. Berlin:* We had to take their word for it; there was no way of looking at every publication.

*Dr. Lucas:* I think we have all accepted today the aphorism that it is unethical to do a badly designed study. Would we also accept the aphorism that it is unethical not to register or make available the results of a clinical trial, regardless of the findings? If we actually regarded it as unethical, then this could be a duty for ethics committees rather than just something that they might or might not take care of; in other words, it should be a specified duty that was regulatory. Is that something that could come out of this meeting?

*Dr. Berlin:* I would say clearly, yes. In fact, we are preparing a chapter for a Cochrane collaboration book, and we managed to get hold of an as yet unpublished manuscript from Iain Chalmers, who was arguing that very thing, that ethics committees are not fulfilling their obligation if they do not insist on registration and publication of results of studies.

**REFERENCE**